

CLAIMS

1. A data processing device for carrying out speech processing in which prediction taps for finding prediction values of the speech of high sound quality are extracted from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, and in which said prediction taps are used along with preset tap coefficients to perform preset predictive calculations to find said prediction values of said speech of high sound quality, said device comprising:

prediction tap extracting means for extracting from said synthesized sound said prediction taps used for predicting said speech of high sound quality, as target speech, the prediction values of which are to be found;

class tap extraction means for extracting a class tap, used for sorting said target speech to one of a plurality of classes, from said code, by way of classification;

classification means for finding the class of said target speech based on said class tap;

acquisition means for acquiring said preset tap coefficients associated with the class of said target speech from among a plurality of tap coefficients as found on learning from class to class; and

prediction means for finding said prediction values of said target speech using said prediction taps and said preset tap coefficients associated with said class of said

target speech.

2. The data processing device according to claim 1 wherein said prediction means perform one-dimensional linear predictive calculations, using said prediction taps and the tap coefficients, to find the prediction values of said target speech.

3. The data processing device according to claim 1 wherein said acquisition means acquires said tap coefficients of the class associated with said target speech from storage means holding said tap coefficients on the class basis.

4. The data processing device according to claim 1 wherein said class tap extraction means extracts said class taps from said code and from said linear prediction coefficients or residual signals obtained on decoding said code.

5. The data processing device according to claim 1 wherein said tap coefficients have been obtained on carrying out learning so that the prediction errors of the predicted values of the speech of high sound quality obtained on carrying out preset predictive calculations employing said prediction taps and said tap coefficients will be statistically minimum.

6. The data processing device according to claim 1 further comprising:
said speech synthesis filter.

7. The data processing device according to claim 1 wherein said code has been obtained on encoding the speech in accordance with the CELP (Code Excited Linear Prediction Coding) system.

8. A data processing method for carrying out speech processing of extracting

prediction taps for finding prediction values of the speech of high sound quality from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, and of performing preset predictive calculations using prediction taps along with preset tap coefficients to find said prediction values of said speech of high sound quality, said method comprising:

a prediction tap extracting step of extracting from said synthesized sound said prediction taps used for predicting said speech of high sound quality, as target speech, the prediction values of which are to be found;

a class tap extraction step of extracting a class tap, used for sorting said target speech to one of a plurality of classes, by way of classification, from said code;

a classification step of finding the class of said target speech based on said class tap;

an acquisition step of acquiring said tap coefficients associated with the class of said target speech from among said tap coefficients as found on learning from class to class; and

a prediction step of finding said prediction values of said target speech using said prediction taps and said tap coefficients associated with said class of said target speech.

9. A recording medium having recorded thereon a program for having a computer

execute speech processing of extracting prediction taps for finding prediction values of the speech of high sound quality from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, and of performing preset predictive calculations using said prediction taps along with preset tap coefficients to find said prediction values of said speech of high sound quality, said method comprising:

a prediction tap extracting step of extracting from said synthesized sound said prediction taps used for predicting said speech of high sound quality, as target speech, the prediction values of which are to be found;

a class tap extraction step of extracting class taps, used for sorting said target speech to one of a plurality of classes, by way of classification, from said code;

a classification step of finding the class of said target speech based on said class taps;

an acquisition step of acquiring said tap coefficients associated with the class of said target speech from among said tap coefficients as found on learning from class to class; and

a prediction step of finding said prediction values of said target speech using said prediction taps and said tap coefficients associated with said class of said target speech.

10. A learning device for learning preset class taps usable for finding, by preset

predictive calculations, prediction values of the speech of high sound quality from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, said learning device comprising:

class tap extraction means for extracting class taps from said code, said class taps being used for classifying said speech of high sound quality, as target speech, the prediction values of which are to be found;

classification means for finding a class of said target speech based on said class taps; and

learning means for carrying out learning so that the prediction errors of the prediction values of the speech of high sound quality obtained on carrying out predictive calculations using said tap coefficients and the synthesized sound will be statistically minimum, to find said tap coefficients from class to class.

11. The learning device according to claim 10 wherein said learning means carries out learning so that the prediction errors of the prediction values of the speech of high sound quality obtained on carrying out one-dimensional linear predictive calculations using said tap coefficients and the synthesized sound will be statistically minimum.

12. The learning device according to claim 10 wherein said class tap extraction means extracts said class taps from said code and from said linear prediction coefficients and said residual signals obtained on decoding said code.

13. The learning device according to claim 10 wherein said code is obtained on encoding the speech in accordance with the CELP (Code Excited Linear Prediction Coding) system.

14. A learning method for learning preset class taps usable for finding, by preset predictive calculations, prediction values of the speech of high sound quality from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, said learning device comprising:

a class tap extraction step of extracting class taps from said code, said class taps being used for classifying said speech of high sound quality, as target speech, the prediction values of which are to be found;

a classification step of finding a class of said target speech based on said class taps; and

a learning step of carrying out learning so that the prediction errors of the prediction values of the speech of high sound quality obtained on carrying out predictive calculations using said tap coefficients and the synthesized sound will be statistically minimum, to find said tap coefficients from class to class.

15. A recording medium having recorded thereon a program for having a computer execute learning processing of learning preset class taps usable for finding, by preset predictive calculations, prediction values of the speech of high sound quality from the

synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, said learning device comprising:

a class tap extraction step of extracting class taps from said code, said class taps being used for classifying said speech of high sound quality, as target speech, the prediction values of which are to be found;

a classification step of finding a class of said target speech based on said class taps; and

a learning step of carrying out learning so that the prediction errors of the prediction values of the speech of high sound quality obtained on carrying out predictive calculations using said tap coefficients and the synthesized sound will be statistically minimum, to find said tap coefficients from class to class.

16. A data processing device for generating, from a preset code, filter data to be afforded to a speech synthesis filter adapted for synthesizing the speech based on linear prediction coefficients and a preset input signal, comprising:

code decoding means for decoding said code to output decoded filter data;

acquisition means for acquiring preset tap coefficients as found by carrying out learning; and

prediction means for carrying out preset predictive calculations, using said tap coefficients and the decoded filter data, to find prediction values of said filter data, to

send the so found prediction values to said speech synthesis filter.

17. The data processing device according to claim 16 wherein said prediction means carries out one-dimensional linear predictive calculations to find prediction values of said filter data.

18. The data processing device according to claim 16 wherein said acquisition means acquires said tap coefficients from storage means holding said tap coefficients.

19. The data processing device according to claim 16 further comprising:

prediction tap extraction means for extracting prediction taps from said decoded filter data, said prediction taps being usable along with said tap coefficients for predicting said filter data, as filter data of interest, the prediction values of which are to be found, said prediction means carrying out predictive calculations using said prediction tap and tap coefficients.

20. The data processing device according to claim 19 further comprising:

class tap extraction means for extracting class taps from said decoded filter data, said class taps being used for sorting said filter data of interest to one of a plurality of classes, by way of classification, and classification means for finding the class for said filter data of interest, based on said class taps;

said prediction means carrying out predictive calculations using said prediction taps and said tap coefficients associated with the class of said filter data of interest.

21. The data processing device according to claim 19 further comprising:

class tap extraction means for extracting class taps from said code, said class

tap being used for sorting said filter data of interest to one of a plurality of classes, by way of classification, and classification means for finding the class for said filter data of interest, based on said class tap;

said prediction means carrying out predictive calculations using said prediction taps and said tap coefficients associated with the class of said filter data of interest.

22. The data processing device according to claim 21 wherein said class tap extraction means extracts said class taps from both said code and said decoded filter data.

23. The data processing device according to claim 16 wherein said tap coefficients have been obtained on carrying out learning so that the prediction errors of the predicted values of said filter data obtained on carrying out preset predictive calculations employing said tap coefficients and decoded filter data will be statistically minimum.

24. The data processing device according to claim 16 wherein said filter data is at least one or both of said input signal and said linear prediction coefficients.

25. The data processing device according to claim 16 further comprising:

said speech synthesis filter.

26. The data processing according to claim 16 wherein said code is obtained on encoding the speech in accordance with the CELP (Code Excited Linear Prediction Coding) system.

27. A data processing method for generating, from a preset code, filter data to be afforded to a speech synthesis filter adapted for synthesizing the speech based on

code decoding means for decoding the code corresponding to filter data to output decoded filter data; and

learning means for carrying out learning so that prediction errors of prediction values of said filter data obtained on carrying out predictive calculations using said tap coefficients and decoded filter data will be statistically smallest to find said tap coefficients.

30. The learning device according to claim 29 wherein said learning means performs the learning so that the prediction errors of the prediction values of said filter data obtained on carrying out one-dimensional linear predictive calculations using said tap coefficients and the decoded filter data will be statistically smallest.

31. The learning device according to claim 29 further comprising:

predictive tap extraction means for extracting from said decoded filter data prediction taps used along with said tap coefficients for predicting said filter data, the prediction values of which are to be found, as said filter data of interest;

said learning means effecting learning so that the prediction errors of prediction values of said filter data obtained on carrying out predictive calculations using said prediction taps and tap coefficients will be statistically smallest.

32. The learning device according to claim 31 further comprising:

class tap extraction means for extracting a class tap from said decoded filter data, said class tap being used for sorting said filter data of interest to one of a plurality of classes, by way of classification, and classification means for finding the

class for said filter data of interest, based on said class tap;

said learning means performing learning so that the prediction errors of prediction values of said filter data obtained on carrying out predictive calculations using said prediction taps and said tap coefficients associated with the class of said filter data of interest will be statistically smallest.

33. The learning device according to claim 31 further comprising:

class tap extraction means for extracting a class tap from said code, said class tap being used for sorting said filter data of interest to one of a plurality of classes, by way of classification, and classification means for finding the class for said filter data of interest, based on said class taps;

said learning means performing learning so that the prediction errors of prediction values of said filter data obtained on carrying out predictive calculations using said prediction taps and tap coefficients will be statistically smallest.

34. The learning device according to claim 33 wherein said class tap extraction means extracts said class taps from both said code and said decoded filter data.

35. The learning device according to claim 29 wherein said filter data is at least one or both of said input signal and said linear prediction coefficients.

36. The learning device according to claim 29 wherein said code is obtained on encoding the speech in accordance with the CELP (Code Excited Linear Prediction Coding) system.

37. A learning method for learning preset tap coefficients usable for finding, by

predictive calculations from a code associated with filter data to be applied to a speech synthesis filter which synthesizes the speech based on linear prediction coefficients and a preset input signal, prediction values of said filter data, comprising:

a code decoding step of decoding the code corresponding to filter data to output decoded filter data; and

a learning step of carrying out learning so that the prediction errors of prediction values of said filter data obtained on carrying out predictive calculations using said tap coefficients and decoded filter data will be statistically smallest to find said tap coefficients.

38. A recording medium having recorded thereon a program for having a computer execute learning processing of learning preset tap coefficients usable for finding, by predictive calculations from a code associated with filter data to be applied to a speech synthesis filter which synthesizes the speech based on linear prediction coefficients and a preset input signal, prediction values of said filter data, comprising:

a code decoding step of decoding the code corresponding to filter data to output decoded filter data; and

a learning step of carrying out learning so that the prediction errors of prediction values of said filter data obtained on carrying out predictive calculations using said tap coefficients and decoded filter data will be statistically smallest to find said tap coefficients.

39. A speech processing device for finding prediction values of the speech of high

sound quality from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, comprising:

prediction tap extraction means for extracting prediction taps usable for predicting the speech of high sound quality, as target speech, the prediction values of which are to be found,

class tap extraction means for extracting class taps, usable for sorting the target speech to one of a plurality of classes, by way of classification, from said synthesized sound, said code or the information derived from said code;

acquisition means for acquiring said tap coefficients associated with the class of said target speech from the tap coefficients as found on learning from one class to another; and

prediction means for finding the prediction values of said target speech using said prediction taps and said tap coefficients associated with the class of said target speech.

40. The data processing device according to claim 39 wherein said prediction means effects one-dimensional linear predictive calculations, using said prediction taps and tap coefficients, to find prediction values of said target speech.

41. The data processing device according to claim 39 wherein said acquisition means acquires said tap coefficients of the class associated with said target speech from

storage means holding said tap coefficients from class to class.

42. The data processing device according to claim 39 wherein said prediction tap extraction means or class tap extraction means extracts said prediction taps or class tap from said synthesized sound, said code or the information derived from said code.

43. The data processing device according to claim 39 wherein said tap coefficients have been obtained on carrying out learning so that the prediction errors of the predicted values of said speech of high sound quality obtained on carrying out preset predictive calculations employing said prediction taps and tap coefficients will be statistically minimum.

44. The data processing device according to claim 39 further comprising:
a speech synthesis filter.

45. The data processing device according to claim 39 wherein said code has been obtained on coding the speech with CELP (Code Excited Linear Prediction Coding) system.

46. A speech processing method for finding prediction values of the speech of high sound quality from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, comprising:

a prediction tap extraction step of extracting prediction taps usable for predicting the speech of high sound quality, as target speech, the prediction values of

a class tap extraction step of extracting class taps, usable for sorting the target speech to one of a plurality of classes, by way of classification, from said synthesized sound, said code or the information derived from said code;

an acquisition step of acquiring said tap coefficients associated with the class of said target speech from the tap coefficients as found on learning from one class to another; and

a prediction step of finding the prediction values of said target speech using said prediction taps and said tap coefficients associated with the class of said target speech.

48. A learning device for learning preset tap coefficients usable for finding, by preset predictive calculations, prediction values of the speech of high sound quality, from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, comprising:

prediction tap extraction means for extracting prediction taps usable in predicting the speech of high sound quality, as target speech, the prediction values of which are to be found, from said synthesized sound, said code or the information derived from said code;

class tap extraction means for extracting class taps usable for sorting the target speech to one of a plurality of classes, by way of classification, from said synthesized sound, said code or the information derived from said code;

classification means for finding the class of said target speech based on said

class taps; and

learning means for carrying out learning so that the prediction errors of prediction values of said speech of high sound quality, obtained on carrying out predictive calculations using said tap coefficients and said prediction taps, will be statistically smallest.

49. The learning device according to claim 48 wherein said learning means carries out learning so that the prediction errors of the prediction values of the speech of high sound quality obtained on carrying out the one-dimensional linear predictive calculations using said tap coefficients and the prediction taps will be statistically smallest.

50. The learning device according to claim 48 wherein said prediction tap extraction means or class tap extraction means extract said prediction taps or the class taps from the synthesized sound, said code and the information derived from said code.

51. The learning device according to claim 48 wherein said code has been obtained on coding the speech with CELP (Code Excited Linear Prediction Coding) system.

52. A learning method for learning preset tap coefficients usable for finding, by preset predictive calculations, prediction values of the speech of high sound quality, from the synthesized sound obtained on affording linear prediction coefficients and residual signals, generated from a preset code, to a speech synthesis filter, said speech of high sound quality being higher in sound quality than said synthesized sound, comprising:

a prediction tap extraction step of extracting prediction taps usable in predicting

to be found, from said synthesized sound, said code or the information derived from said code;

a class tap extraction step of extracting a class tap usable for sorting the target speech to one of a plurality of classes, by way of classification, from said synthesized sound, said code or the information derived from said code;

a classification step of finding the class of said target speech based on said class tap; and

a learning step of carrying out learning so that the prediction errors of prediction values of said speech of high sound quality, obtained on carrying out predictive calculations using said tap coefficients and said prediction taps, will be statistically smallest, to find said tap coefficients.